Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2004 Proceedings

Americas Conference on Information Systems (AMCIS)

12-31-2004

Using Web Analytics to Measure the Activity in a Research-Oriented Online Community

Catherine Dwyer
Pace University

Yi Zhang
New Jersey Institute of Technology

Starr Roxanne Hiltz
New Jersey Institute of Technology, roxanne.hiltz@gmail.com

Using Web Analytics to Measure the Activity in a Research-Oriented Online Community

Catherine Dwyer
Pace University
cdwyer@pace.edu

Yi Zhang
New Jersey Institute of Technology
zhang.yi@njit.edu

Starr Roxanne Hiltz
New Jersey Institute of Technology
hiltz@njit.edu

ABSTRACT

Web-based online communities, which require tremendous effort to create and maintain, can nevertheless appear quite inactive. This leads to difficulties justifying their continued support. In our research, we discovered a significant amount of member activity in a seemingly passive online community. We want to argue this activity demonstrates a clear engagement and participation of members in the community. To do this, we will apply Web analytics to a research oriented virtual community and measure member's usage characteristics and interaction with the site. Web analytics uses data mining techniques to uncover browsing patterns by examining the content of Web server logs. It has been successfully applied in usability studies to highlight Web design problems, and has also been used in e-commerce to predict and/or influence shopper's choices. Our application of Web analytics shows that although this community does not appear to have many "voices" in the form of frequent discussion postings, there is significant evidence of both social engagement and information seeking activity by a large number of members on a regular basis.

Keywords

Web Analytics, Online Community, User Online Behavior, Social Engagement, Information Seeking, Web Metrics.

INTRODUCTION

Online communities are very popular social structures because they bring together people with similar interests. In contrast to geographically based communities, online communities require computer systems to support their existence. Various computer systems can support online communities: listservs (mailing lists), Usenet news groups, bulletin board systems, Multiple User Dungeons (MUDs), Multiple-user, Object-Oriented environments (MOOs), chat rooms, virtual reality systems, Web bloggers, and Web sites.

Online research communities usually use listserv or Web sites as their virtual settlement. Listserv-based communities use e-mail as transfer media, thus members' participation is fundamentally interactive and visible. On the other hand, listserv information is delivered in a sequential form and topics are spontaneous, so these sites are incomplete and somewhat disorganized. In contrast, Web-based communities are fundamentally structured and encyclopedic, but members' participation is not as interactive or as visible. Members are also reluctant to contribute ideas because the site is more public and formal than e-mail. But does this mean members are not using the community? It follows that more work needs to be done to capture members' invisible engagement in communities hosted on Web sites.

In this paper, we classify member engagement into two categories: information seeking and social engagement. The selection of these two categories is based on previous work that describes the crucial nature of these two behaviors (Weick 1979, Kiesler et al. 1985; Rockart & DeLong 1988, Preece 2000, Rheingold 1993).

Information seeking is the primary goal for members of online communities, especially research-oriented communities. Therefore, a fundamental objective of research communities is to gather, update, and share all information relating to the community's general interests. Thus, member loyalty will depend on the quality, organization, presentation, and timeliness of the information made available in the community.

Social engagement is another reason why members participate in an online community. They want to meet other people, seek help, support, friendship, love, etc. In other words, they are driven to develop social relationships within the community. Interpersonal communication contributes structure within the community (Weick 1979). Thus, support for interpersonal relationships and encouragement of social engagement are also crucial objectives for a successful online community. Social engagement directly affects the satisfaction of its members. Member satisfaction is a potent predictor of their continued participation in the online community, and it is also a crucial factor for the continued use of the system (Kiesler et al. 1985; Rockart & DeLong 1988). For an online research community, an additional professional and social goal is to establish contact and collaborations with other community members. This requires that members of the community are aware of one another, allowing opportunities to collaborate with other members (Kraut et al. 1990).

Previous research judged whether an online community was successful or not by looking at visible activities, such as postings and forum messages (Girgensohn and Lee 2002, Millen and Patterson 2002). We argue this is not sufficient to completely evaluate Web-based communities, because browsing behavior is an invisible but fundamental motivation for their visitors. Lurkers, those who rarely if ever post yet follow the group's discussions regularly (Jargon Dictionary 2001), are the silent majority in most online communities. Lurkers are reported to make up over 90% of online communities (Mason 1999, Katz 1998). Although lurkers are mostly silent in a community, they are still a very important portion of a community, and their participation should be reflected in the overall evaluation of an online community. Different community systems and different types of communities will have different levels of lurking. For example, health-support discussion lists (DLs) have on average significantly fewer lurkers (46%) than software-support DLs (82%) (Nonnecke and Preece 2001). In order to capture a more complete picture of both active members and lurkers, we applied "Web analytics" to both discover and measure online engagement for Web-based online communities that captures both visible and invisible activities.

This paper is organized as follows: in the next section, we describe the basic methodology for Web Analytics. We then describe how the Web log data were collected and organized. We next examine the data collected to discover patterns and usage trends. Following the examination, we analyze the patterns discovered and draw conclusions on the results. And finally we present our findings, contributions, and potential for future research.

WEB ANALYTICS

Web analytics, which uses Web server logs, is a well established method in e-commerce and Web usability research to uncover and understand online behavior. The use of Web analytics in this application will allow us to obtain a more complete measure of member activities both visible and invisible through the use of entries from server logs. This analysis of real behavior shows a more accurate measure of commitment of members to the community. This documentation of actual behavior is more credible evidence of the community's success in meeting its purpose and overall objectives.

The study of Web server logs in order to uncover visitor usage patterns has evolved over time through quite a variety of names, such as Web traffic monitoring, clickstream analysis, Web e-metrics, and Web usage mining. Each of these terms generally refers to "the application of data mining techniques to discover usage patterns from Web data" (Srivastava, Cooley, Deshpande, and Tan, 2000).

In order to fully explain how this data analysis takes place, we must first define some basic terms (World Wide Web Consortium, 1999):

- **hit** –a server request for a physical file. This is often a misleading indicator of Web activity, since the hit could be the request for an image, frame, JavaScript module, or Web page
- **page view** – the combination of every file or element that contributes to a browser display at one time. It is sometimes referred to as a "logical page," and it is essentially what you see when you look at a Web site
- **clickstream** – sequential series of page view requests
- **visitor** – single individual accessing files from one or more Web servers through a browser
- **cookie** – data sent by a Web server to a Web client, to be stored locally by the client and sent back to the server on subsequent requests
- **server session** – clickstream of page views for a single visitor for a particular Web site, also known as a visit

The primary input data for most applications of Web analytics comes from Web server logs. For example, commercially available Web analyzer products such as Web Trends (http://www.webtrends.com/) use these logs to produce daily and monthly activity reports.

Using server logs to analyze visitor activity can be problematic. A Web server log is simply a log of all requests made to a server. Its fundamental purpose is to capture information useful to technicians managing and maintaining a Web server, so the data collected are not in a format easily used to model visitor activity. For example, every hit creates a log entry. Therefore an accurate representation of relevant page views requires recognizing, collecting, and organizing relevant hits in order to re-construct page views. Also, the existence of "spiders" that crawl the Web to capture content for search engines like Google can artificially inflate the number of hits a server receives, and these must be weeded out (Cooley, Mobasher and Srivastava, 1999; Mobasher, Cooley and Srivastava, 2000; Montgomery and Faloutsos, 2000). This process has been quite accurately described as "drinking from the fire hose" (Cooley, 2004).

The basic application of Web analytics requires these three steps:

- Pre-processing: conversion of log input into abstractions and useful formats
- Pattern discovery: application of statistical analysis, association, clustering, and classification
- Pattern analysis: focus on meaningful results and filter out insignificant results

The virtual community that is the subject of our study is the "Asynchronous Learning Networks (ALN) Research Community." The overall objective of this community is to increase the quality, quantity, dissemination, and application of results of research on the effectiveness of ALN, a form of e-learning that combines self-study with substantial, rapid, asynchronous interactivity with others (http://www.aln.org). The community is hosted at http://www.alnresearch.org. Further details about the ALN Research community can be found in (Zhang, Bot and Hiltz, 2001).

An initial analysis of the visible behavior of this community showed members seemed to be inactive and passive. There were few instances of posted forum messages or member contributions of their own research articles to the site. But we argue visible activity does not capture the complete picture, so in combination with the evidence in Web logs, we can more accurately pull out the passive activity to see meaningful measures and determine the true degree of visitor activity. In the following three sections, we are going to apply the Web analytics to the ALN research community web logs to discover real usage in this community.

Pre-processing

In order to convert log data into a manageable format, the first step in Web usage mining is always "data cleaning," (Heer and Chi 2002). For example, the multiple hits generating a single page view need to be consolidated. While the techniques for using Web server logs are well documented (Cooley, et. al. 1999; Pirolli and Pitkow 1999), for the purpose of this study the server software for the ALN Web Center was adapted to collect relevant information and log each page view delivered by the server. This greatly simplified our data cleaning step, since we no longer had to reconstruct page views from their composite hits.

Members of the ALN Web center are required to log in before they can access content. Therefore the custom log can include member id information, in order to allow reconstructing each member's server session. The ability to follow member activity will enable a longitudinal study of changes in their activity as new features and changes are added to the site. It will also allow us to correlate member survey results and demographic information with records of clickstream behavior.

For the purpose of this study we examined web server data from a period of November. 1, 2003 to November 30, 2003.The custom log produces a text file for each calendar date, i.e. 111103.log. Each log entry summarizes a page view requested by a member. The entry includes a timestamp, that session's cookie, the start time of the session (in milliseconds), the time of this page view request (in milliseconds), user identifier (available from required login), page view description, and (if applicable) an article identifier, a number that indicates the visitor has clicked on a link for a stored resource or document, such as a paper abstract or full text.

Below is an example of activity logged from November 11, 2003 in the original server logs. The user logins are changed to preserve the anonymity of the members (see Figure 1):

```
10:44 f0301641161068561365469 1068561365422 1068565470391 visitor01 resources_research_instruments
10:48 f0301641161068561365469 1068561365422 1068565691359 visitor01 resources_learning_theories
10:49 f0301641161068561365469 1068561365422 1068565756172 visitor01 resources_methods
10:50 f0301641161068561365469 1068561365422 1068565804281 visitor01 resources_journals
10:52 f0301641161068561365469 1068561365422 1068565964187 visitor01 resources_evaluation
10:54 f03082993106856606 7016 visitor02 NewSession******************************
10:54 f0308299310685660 67016 1068566066984 1068566089328 visitor02 paper_db_entry 31
11:16 f0301084851068567392016 visitor03 NewSession******************************
11:16 f0301084851068567392016 1068567391984 1068567411469 visitor03 home
11:17 f0301084851068567392016 1068567391984 1068567426406 visitor03  paper_abstract 158
11:18 f0301084851068567392016 1068567391984 1068567518719 visitor03 paper_abstract 142
11:19 f0301084851068567392016 1068567391984 1068567579703 visitor03 paper_abstract 148
11:41 f0301503611068568545328 visitor04 NewSession******************************
11:41 f0301503611068568545328 1068568545266 1068568918891 visitor04 resources_research_instruments
11:43 f0301503611068568545328 1068568545266 1068569016156 visitor04 resources_learning_theories
11:44 f0301503611068568545328 1068568545266 1068569047406 visitor04 project_people
11:44 f0301503611068568545328 1068568545266 1068569085328 visitor04 project_people_proposal
11:46 f0301934121068569210047 visitor05 NewSession******************************
11:46 f0301934121068569210047 1068569209984 1068569217609 visitor05 home
```

Figure 1. Sample raw data from Novemeber 11, 2003

This raw information was consolidated then input into a pre-processing program. The program collected entries by session (identified by its cookie), then recreated the clickstream for each server session recorded.

Table 1 shows a sample clickstream of a server session from the pre-processing for the same date (see Table 1):

Page View #	Start time (HH:MM)	Page View Description	Article ID
1	9:45	NewSession******************************	
2	9:45	home	
3	9:48	paper_abstract	158
4	9:57	paper_full_text	158
5	10:14	paper_full_text	158
6	10:14	paper_abstract	158
7	10:15	paper_abstract	157
8	10:16	paper_full_text	157
9	10:21	paper_abstract	154
10	10:22	paper_abstract	153
11	10:22	paper_abstract	153

12	10:25	paper_full_text	153
13	10:26	paper_abstract	148
14	10:27	paper_full_text	148
15	10:28	paper_abstract	67
16	10:29	paper_full_text	67
17	10:43	people_directory	
18	10:43	project_people	
19	10:43	forum_list	
20	10:44	resources_research_instruments	
21	10:48	resources_learning_theories	
22	10:49	resources_methods	
23	10:50	resources_journals	
24	10:52	resources_evaluation	

Table 1: Sample Clickstream of a server session

Table 1 summarizes the clickstream path taken by a visitor. It began at 9:45 am, and its last page view was at 10:52 am. The session lasted one hour, and seven minutes. It consisted of 24 different page views. This visitor looked at related entries (i.e. abstract and full text) for six different papers. For example we see the visitor viewing the abstract then the full text in entries 3 and 4, 13 and 14, and 15 and 16. The visitor also viewed information about other ALN members (page view #17), information about the directors of the project (page view #18), and briefly looked at other resources in the site, such as research instruments (#20) and learning theories (#21). Based on a simple analysis of just one session, we can start to see patterns in this member's interaction with a virtual community. Spending over an hour at a Web site is quite a significant amount of time, and during that interval he or she conducted a fairly extensive survey of the content and resources available at the Web center.

From the pre-processing, we also obtain a table of session summaries. Table 2 shows a subset of the session summaries of the same day:

Session #:	Visitor:	Start time:	Total time:	# of pages viewed:	# articles accessed:
1	Visitor1	2:21	0:05	36	0
2	Visitor2	2:26	0:00	2	0
3	Visitor3	2:43	0:20	48	6
4	Visitor4	2:47	0:09	15	8
5	Visitor5	3:09	0:11	10	0
6	Visitor6	3:37	0:05	23	4
7	Visitor7	3:48	0:18	10	0
8	Visitor8	4:26	0:06	11	1
9	Visitor9	4:28	0:33	9	2
10	Visitor10	8:46	0:17	46	15
11	Visitor11	9:14	0:15	36	4
12	Visitor12	10:00	0:12	25	19
13	Visitor12	10:24	0:00	2	1
14	Visitor13	12:27	0:06	28	0
15	Visitor14	12:36	10:18	49	40
16	Visitor15	12:41	0:05	22	7

Table 2: Sample Session Summaries

We can see that one user (number 15) appears to have spent just about all day at the community site, and accessed 40 articles.

Pattern Discovery

The pattern discovery process derives statistical results from pre-processed data by finding associations among data, and clustering and classifying data. The output of the pre-processing step provides us with important summary information about member visits during the period of study.

	total	average/visitor	average for site/day
Number visitors	154		5.13
Number days visited	30	1.87	
Number page views	9114	56.37	303.8
Number sessions	458	2.97	15.26
Number of articles accessed	2061	13.38	68.7

Table 3: Web center summary 11/1/2003-11/30/2003

By studying the summary data, we can see that the Web Center has a regular stream of visitors that are spending time on a number of resources, and returning to visit. The average visitor initiated 2.97 sessions, visiting 1.87 days during the sample month and downloading 13.38 articles.

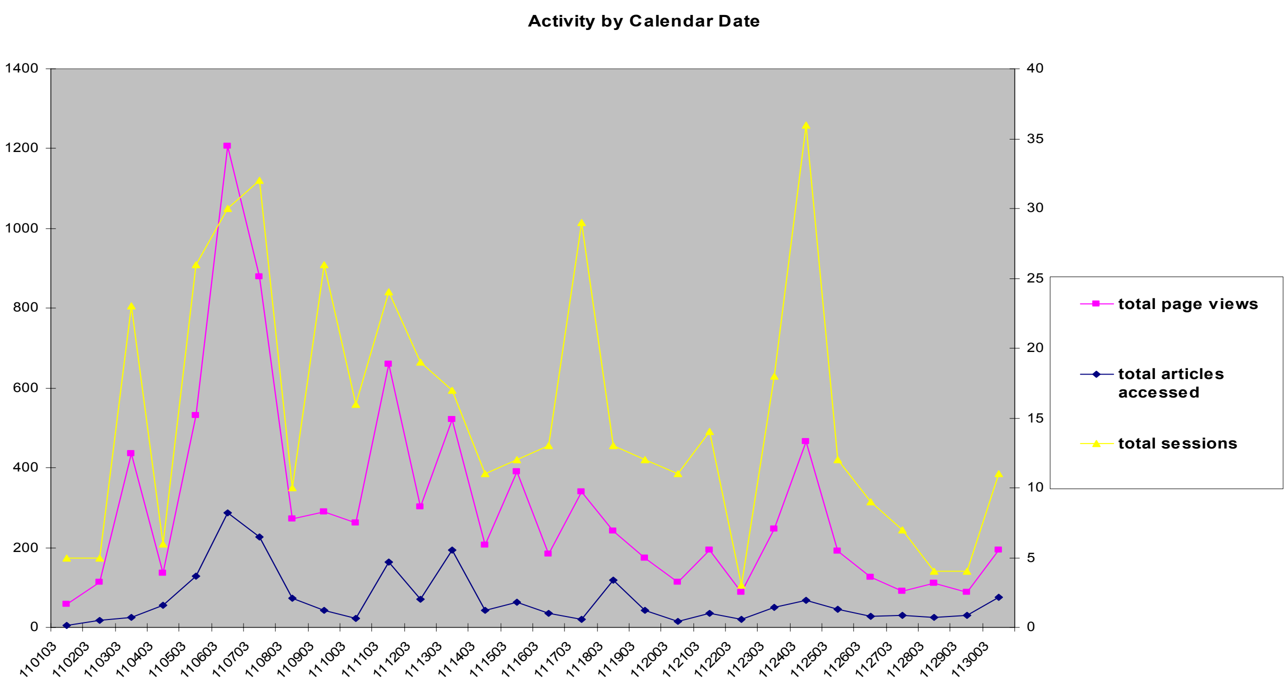

Figure 2. Activity By Calendar Date

Figure 2 describes member activity over the month by tracking the number of articles accessed, the number of sessions, and the number of page views for each date. The chart shows that while there are some noticeable spikes and valleys in the data, the trend is consistent member activity throughout the month (daily number of page views falls between 200 and 600 range). The highest spike, on 11/06, followed an e-mail posting on the ALN Research Center to the ISWorld listserv.

Pattern Analysis

During pattern discovery, we gathered summary data of members' behavior. In pattern analysis, we will classify different activities and carry out more detailed analysis on the findings. The following research questions will guide our pattern analysis process.

- Is there any dominant activity in this community?
- What is the ratio of social engagement activities and information seeking activities?

Dominant activity

The summary data gathered in the previous section showed just over 9,000 page views delivered over a one month period. Based on the structure of the Web site, we further organized these page views into the following categories:

- Project – view information about the project organizers and advisory board
- People directory – view information about other members
- Comments – view and add comments to the site
- Forum discussion forum
- Contributions – input form to contribute a paper to the site
- Tutorials – tutorials on research methods and measuring learning effectiveness
- Resources – instruments useful for further research and studies
- Other Studies – publications following other methodologies
- Empirical Studies – – published papers that follow empirical research methodology

Collecting and counting each page view by category, we can see a dominant usage pattern emerge. Here is a graph demonstrating the distribution of member page views over the course of a week:

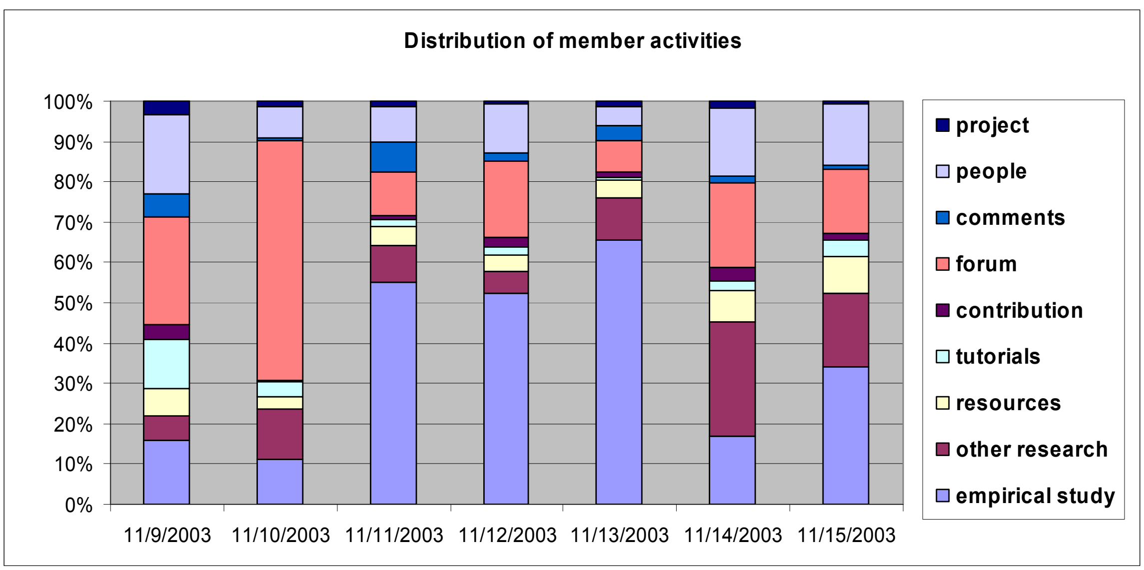

Figure 3. Sample Distribution of Member Activities

From the above graph, we can see that for most days, the dominant activity is viewing empirical research papers (labeled empirical study in the graph). While this chart only shows one week, the results are consistent throughout the month. The most dominant activity for member is the access of empirical research studies.

Although dominant, it is by no means the overwhelming activity day in and day out. In fact, on three of the seven days in this sample, empirical study was *not* the leading activity. Although the amount varies day by day, it is evident that members are accessing all the parts of the web site on a regular basis. This reinforces the pattern beginning to emerge from this server data that members come to the site and thoroughly explore what it has to offer.

An alternative method for measuring the dominant activity is by adding up time spent time spent on each category. However, this method is not as accurate because as we cannot correctly record a member's viewing time on a page – they might be on the phone or using another application. Therefore, we choose to use a count of page views as the criteria to judge the dominant activity.

Ratio of social engagement to information seeking behavior

In order to investigate the ratio of social activities to information activities, we classify all the activities into two types: social engagement and information seeking. Because some activities can be motivated by either social engagement and/or information seeking, the classifying method is not unique. An example of this is a page view that displays a message from a forum. The member could be either looking for information or participating in an asynchronous conversation. So this page view counts as both an information seeking and a social engagement activity.

In order to study the relationship between social engagement and information seeking, all possible page views related to these activities were identified. They were then labeled as either information seeking, social engagement, and in a few cases, put in both categories (see Table 3). Then each page view was assigned a score. The score represents the amount of effort the member would have to engage in to carry out that activity. For example, looking at an abstract is an example of information seeking activity, but looking at the full text is an activity of a higher effort level. Therefore viewing an abstract is assigned a score of one, while viewing the full page paper is assigned a value of two. Here, we need to mention that the way of assigning scores is not unique or standard, we choose it as a rough measure of the time and effort members spend as the they carry out these selected activities.

Category	category	Page_view description	Score
Information Seeking	Empirical Study	paper_abstract	1
		paper_comments_display	1
		paper_db_entry	2
		paper_full_text	2
		paper_new_articles	1
		paper_search_database	1
		published_research	1
	Other Studies	books	1
		books_abstract	1
		other	1
		other_abstract	1
		other_comments_display	1
		other_full_text	2
		other_summary	1
		qualitative	1
		thesis	1
		theories	1
		thesis_abstract	1
	Other Studies	technology	1
	Resources	resources_evaluation	1
		resources_journals	1

		resources_learning_theories	1
		resources_methods	1
		resources_other	1
		resources_research_instruments	2
	Tutorials	tutorials	1
		tutorials_assessment_tutorial	3
	Forum Access	forum_search	1
		forum_search_results	1
		forum_list	1
		forum_messages	1
		forum_replies	1
	Comment Access	paper_comments_display	1
		other_comments_display	1
Social Engagement	Contributions	contribute_instruments	2
		contribute_books	1
		contribute_links	1
		contribute_others	1
		contribute_papers	2
		contribute_studies	3
		contribute_thesis	2
	Message Posting	forum_create_thread	1
		forum_post_message	2
		forum_reply	2
		quote	1
		post_comments	2
	Forum Access	forum_search	1
		forum_search_results	1
		forum_list	1
		forum_messages	1
		forum_replies	1
	Comment Access	paper_comments_display	1
		other_comments_display	1
	People Directory	people_directory	1
	Project	project_people	1
		project_people_proposal	1

Table 4. Page_view Categories

So by congregating page views as either social engagement or information seeking, then adding up the assigned score for each page view, each day was measured as to the amount of social engagement versus information seeking activity that was measured on the site. The following graph shows the pattern that we have discovered (see Figure 4):

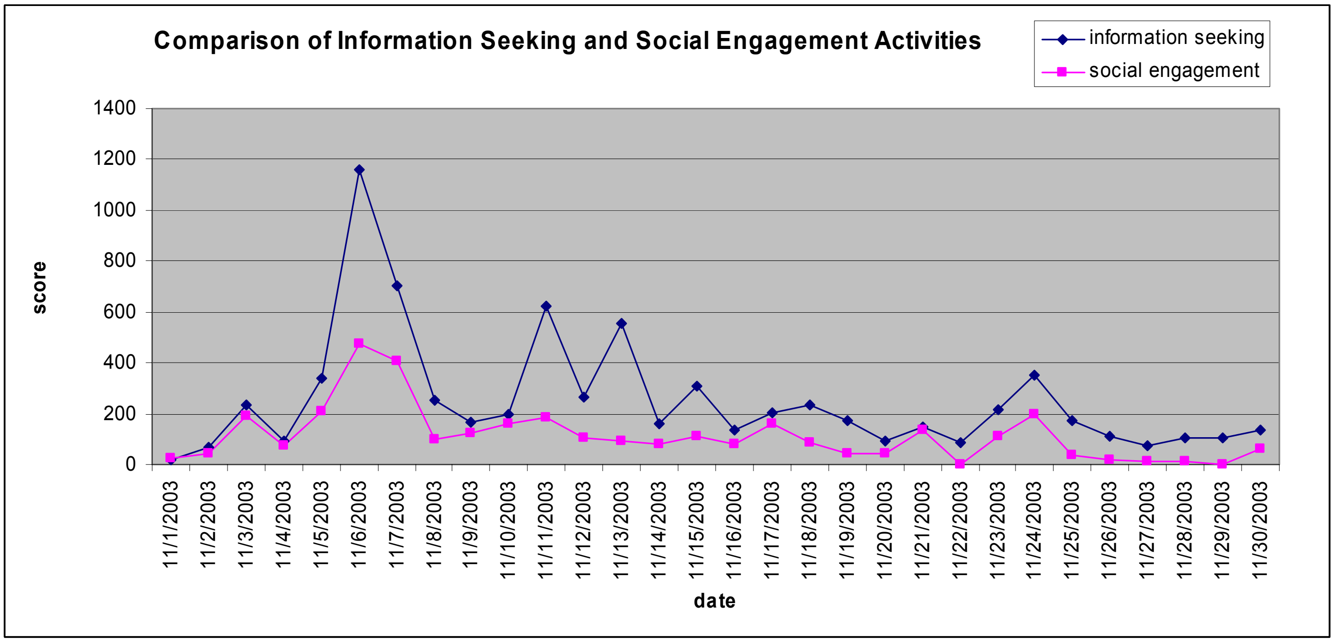

Figure 4. Information Seeking vs. Social Engagement on Sample Dates

An interest effect of this analysis shows that the ratio of social engagement to information seeking moves consistently in relation to each other throughout the month. This clearly demonstrates that despite a lack a visible signs of member engagement, the importance of both categories is apparent from these results. Except for several spike and valley days, the ratio between information seeking and social engagement behavior is almost consistent from 11/1/2003 to 11/30/2003. If you compare the chart of page views contained in Figure 2, you will see a very similar pattern. Notice the similar spikes on the same dates, .

CONCLUSIONS AND FUTURE RESEARCH

The motivation behind our study was to uncover a clearer picture of member activity at Web-based online communities as they are usually fairly inactive in terms of members' postings. This is an especially important goal for research based communities, which are only maintained with enormous effort and continuous influx of grant and institutional funding.

We believe our discovery of invisible member behavior over a one month period shows a significant amount of member activity at a seemingly inactive site. And in addition, this activity contained both information seeking behavior and social engagement activity, and this ratio was quite consistent throughout the month, although this will need to be confirmed with a larger sample size. This leads to the following themes for future research: what is a "good" ratio, and what events cause the change of the ratio.

Our contribution is the demonstration of how invisible activities by "lurkers" can present a more accurate evaluation of a web-based community, and an explanation of how to apply the technique known as Web analytics to the study of other kinds of virtual communities.

There is much potential for future research from our results. We have continued to accumulate server data, and can now undertake a longer term study of both over Web Center usage and the pattern over time of individual members. We also plan to use this data to measure the effect on overall and individual usage by the addition of personalization tools. Now that we know we can capture both overall and individual usage patterns, we can carry out further studies to determine whether demographic factors influence clickstream patterns or overall social engagement versus information seeking behavior. And finally, our results clearly show that the WebCenter is fulfilling it primary mission to "increase the quality, quantity, and dissemination of results of research on the effectiveness of Asynchronous Learning Networks (ALN)."

ACKNOWLEDGEMENTS

This research was partially supported by the Alfred P. Sloan Foundation and the National Science Foundation (EISA-9818309). The opinions expressed are solely those of the authors.

REFERENCES

1. Cooley, R., Mobasher, B., and Srivastava, J. (1999) Data preparation for mining world wide Web browsing patterns, *Knowledge and Information Systems* 1(1).
2. Cooley, R. (2004) Drinking from the Firehose: Converting Raw Web Traffic and E-Commerce Data Streams for Data Mining and Marketing Analysis, available at http://www.Webusagemining.com/sys-tmpl/Webdataminingworkshop/.
3. Girgensohn, A., & Lee, A. (2002). Making Web Sites Be Places for Social Interaction*. Paper presented at the Computer Supported Collaborative Work'02*, New Orleans, Louisiana.
4. Heer, J. and Chi, E. (2002) Separating the Swarm: Categorization Methods for User Sessions on the Web, Proceedings of *CHI 2002 Conference*, ACM Press, 243-250.
5. Katz, J. Luring the Lurkers. Available at http://www.netmeg.net/jargon
6. Kiesler, S., Zubrow, D., Moses, A. M., & Geller, V. (1985). Affection in computer-mediated communication. *Human Computer Interaction, 1*, pp. 77-104.
7. Mason, B. (1999). Issues in Virtual Ethnograph published in Ethnographic Studies in Real and Virtual Environments: Inhabited Information Spaces and Connected Communities. *Paper presented at the Esprit i3 Workshop on Ethnographic Studies*. Edinburgh, 61-69.
8. Millen, D. R., & Patterson, J. F. (2002). *All ways aware: Simulating social engagement in a community network.* Paper presented at the 2002 ACM conference on computer supported cooperative work, New Orleans, Louisiana, 306-313.
9. Mobasher, B., Cooley, R., and Srivastava, J. (2000) Automatic Personalization Based on Web Usage Mining, *Communications of the ACM*, 43, 8.
10. Montgomery, A. and Faloutsos, C. (2001) Identifying Web Browsing Trends and Patterns, *IEEE Computer*, 34, 7.
11. Nonnecke, B., & Preece, J. (2001). *Why Lurkers Lurk.* Paper presented at the Americas Conference on Information Systems, Bostom, pp. 1520-1530.
12. Pirolli, P., Pitkow, L., James E. (1999) Distributions of surfers' paths through the World Wide Web: Empirical characterization. *World Wide Web*, 2, 29-45.
13. Preece, J. (2000). Online Communities: Designing Usability, Supporting Sociability. Chichester, UK: John Wiley & Sons.
14. Rheingold, H. (1993). *The Virtual Community*: MIT Press.
15. Rockart, J. F., & DeLong, D. W. (1988). *Executive support systems: The emergence of top management computer use*. Homewood, IL: Dow Jones-Irwin.
16. Srivastava, J. , Robert C., Mukund D., and Tan, P. (2000) Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, *SIGKDD Explorations*, 2000.
17. Weick, K. E. (1979). *The Social Psychology of Organizing*. Reading, MA: Addison-Wesley.
18. World Wide Web Consortium (1999) Web Characterization Activity, available at http://www.w3.org/WCA.
19. Zhang, Y., Bot, R., and Hiltz, S. R. (2001) Building a Virtual Community for ALN, *Proceedings of the Seventh American Conference on Information Systems*, Boston.